Sources of variation in forced expiratory volume in one second and forced vital capacity

S. Chinn*, D. Jarvis*, C. Svanes# and P. Burney*

ABSTRACT: Published prediction equations for lung function differ considerably, but the components of population variation responsible for the differences are unknown.

Data were analysed for 6,323 never-smoking adults who did not report wheeze or asthma, from 42 centres participating in the European Community Respiratory Health Survey. Means and components of variance were estimated for males and females aged 20–24 yrs, and the relationships with age and height were examined in those aged 25–44 yrs.

Mean lung function for those aged 20–24 yrs differed between centres, but variation could not be wholly attributed to differences in population or equipment. The maximum difference in means by equipment type was 101 mL for FVC in males. Equipment differences were not statistically significant adjusted for country, but differences in mean forced expiratory volume in one second and forced vital capacity by country, adjusted for instrument, were statistically significant in males. Differences between centres in relation to age and height had less influence on predicted values.

In conclusion, there are unexplained differences in lung function between ethnically similar nonsmoking symptom-free populations. Neither national reference curves nor those based on the same ethnic group can be guaranteed to give accurate norms of lung health.

KEYWORDS: Height, populations, reference values, spirometry

AFFILIATIONS
*Dept of Public Health Sciences, King's College London, London, UK.
#Dept of Thoracic Medicine, Haukeland Hospital, Bergen, Norway.

CORRESPONDENCE
S. Chinn
Dept of Public Health Sciences
King's College London
5th Floor Capital House
42 Weston Street
London SE1 3QD
UK
Fax: 44 2078486605
E-mail: sue.chinn@kcl.ac.uk

Spirometric lung function measurements are used clinically for diagnosis and monitoring, and have many uses in research. In clinical use, forced expiratory volume in one second (FEV1) and forced vital capacity (FVC) are usually each expressed as a percentage of predicted value for height, age and sex. For example, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria classify patients with chronic obstructive pulmonary disease into categories according to post-bronchodilator FEV1/FVC and FEV1 % predicted [1]. Researchers often analyse FEV1 % pred as an outcome variable [2, 3] or select patient groups according to FEV1 % pred [4].

The GOLD criteria do not specify the reference equations that should be used to obtain predicted values, despite evidence that the published predicted values for FEV1 and FVC differ considerably [5], for both males and females. Recent recommendations for lung function testing have suggested that reference values should be derived from lung function measurements from a population including the "age range, sex and ethnic group of individuals to be tested" [6]. This assumes that predicted values should depend only on these three parameters. Each of the equations compared by ROCA et al. [5] was a regression equation linear in height and age fitted to data for adults, mostly from those aged 25 yrs [7–11]. Hence, reference equations that are usually constructed using data from adults have four elements: the mean value at the youngest age considered fully mature; the adjustment for height; the relationship with age; and the residual standard deviation. Each of these may differ between populations and over time.

The reasons for the differences in predicted values were not explored by ROCA et al. [5], but the regression coefficients in the published equations suggest that these include variation in the relationship with height or age between populations. Calculations show some variation in predicted values for those aged 25 yrs and those of average height.

ROCA et al. [5] found variation between the mean values for adults aged 20–44 yrs from 34 centres participating in the European Community

Respiratory Health Survey (ECRHS). The variation was not partitioned between the mean values in young adults, or different relationships to age and height, the knowledge of which can inform debate about appropriate reference values. The analysis presented in the current study describes variation in FEV1 and FVC in healthy nonsmoking adults aged 20–24 yrs in 42 centres taking part in the ECRHS I. The relationships with height and age in adults aged 25–44 yrs are also estimated.

METHODS

Participants and study design

The protocol for ECRHS I has been described in detail elsewhere [12]. Participating centres invited young adults from an area defined by pre-existing administrative boundaries, with a population of $\geqslant$150,000 people. At stage one, where possible, an up-to-date sampling frame was used to randomly select at least 1,500 males and 1,500 females aged 20–44 yrs, who were sent a self-completed postal questionnaire. A random sample of responders to stage one was invited to stage two, which included an administered questionnaire and measurement of lung function. The administered questionnaire in stage two included the question: ''Have you ever smoked for as long as a year?'' Stage two was carried out from 1990 to 1995 across 42 centres, which included the 34 centres in the analysis of ROCA et al. [5], and eight contributing data later. The centre in Bombay (India) was omitted, as were those in Aarhus (Denmark) and Wroclaw (Poland) where the equipment used was unknown. Most centres were in Western Europe, plus three in New Zealand, one in Australia, six in Canada and one in the USA. Ethnic origin was not recorded, but participants were known to be almost exclusively White.

Spirometry

The equipment used was Biomedin spirometer (Biomedin, Padua, Italy; 16 centres), SensorMedics spirometer (SensorMedics, Yorba Linda, CA, USA; nine centres), Spirotech spirometer (Spirotech, Bilthoven, the Netherlands; eight centres), Jaeger pneumotach (Jaeger, Hoechberg, Germany; three centres), Morgan spirometer (Morgan, Haverhill, MA, USA; three centres), Morgan pneumotach (Morgan; one centre), Fleisch pneumotach connected to a Hewlett-Packers lung function analyser (Massach USA; one centre) and Vitalograph spirometer (Vitalograph, Buckingham, UK; one centre), each of which complied with American Thoracic Society (ATS) standards. The maximum FEV1 and maximum FVC of up to five technically acceptable blows were determined, and whether FEV1 and FVC each met the ATS criterion for reproducibility [13]. Height was recorded prior to spirometry. Out of the 42 centres, it was measured in 31, self-reported in five centres, and not recorded whether measured or asked in six. Each European centre took part in a centrally organised training day, and was visited by a central organiser who checked that the common protocol for lung function testing was observed.

Statistical analysis

All analyses were carried out for never-smokers who did not report ever having had asthma, or wheeze in the last 12 months, for males and females separately. Data were divided

a priori between participants aged 20–24 yrs and those aged 25–44 yrs, since 25 yrs has been reported as the age from which lung function starts to decline [11], at least in males [8]. FALASCHETTI et al. [14] demonstrated a plateau up to around age 25 yrs for FEV1 and FVC in males, and for FVC in females. Variation of FEV1 and FVC in those aged 20–24 yrs was analysed for heterogeneity between centres using Bartlett's test. The variation between individuals and between centres was estimated without and with adjustment for height. This was carried out using multilevel models with participant at level one and centre at level two, with height included as a covariate at level one for the height-adjusted components. The regression coefficients for height from the models were used to calculate height-adjusted values. Values were analysed according to whether they met the ATS reproducibility criterion or not, i.e. that the two largest values did not differ by >0.2 L [13]. Height-adjusted centre means were analysed in relation to response rate, type of instrument used and country. A meta-analysis method was used to estimate heterogeneity of means and relationships with height and age between centres, for those aged 25–44 yrs [15]. The percentage of total variation across centres due to chance was calculated [16].

RESULTS

In the 42 centres, there were 44,801 respondents to the stage one questionnaire in the sample randomly selected for stage two, of whom 21,503 (48.0%) participated in stage two and 18,160 (40.5%) had an acceptable FEV1. Response rates varied considerably between centres (table 1). Of those taking part in stage two, 18,160 (85.5%) provided an acceptable FEV1, ranging 52.2–98.5% across centres. Of these, 17,545 (96.6%) met the ATS criterion. The number of nonsmokers who did not report wheeze in the last 12 months or ever having asthma was 6,323, of whom 1,234 were aged <25 yrs (table 1).

Adults aged 20–24 yrs

There were 605 males aged <25 yrs who did not report wheeze, ever having asthma or ever having smoked for as long as 1 yr with a valid FEV1. Of these, 604 also had a valid FVC. There were 629 females meeting the criterion with FEV1 and FVC measurements. There was some evidence that the degree of variation in FEV1 (p=0.042 for males; p=0.047 for females) and in FVC for females (p=0.041) differed between centres, but this was not the case for FVC in males (p=0.85).

Table 2 shows the mean, average within-centre variation and between-centre variation for males and females. Without adjustment for height, differences between centre means accounted for, at most, 10% of the between-person variation, as shown by the intraclass correlation coefficients (ICCs). This centre variation increased the standard deviation of a single lung function measurement by at least 3%, for females, but at most 6%, for FEV1 in males, as shown by the ratios of the total single-determination standard deviation to the within-centre standard deviation in table 2. No statistically significant variation between centres in the relationship of FEV1 or FVC with height was detected, either in males or females. The mean$\pm$SD height of the young males was 1.79$\pm$0.07 m and 1.66$\pm$0.07 m for the young females. Adjustment for height reduced each component of variation, but had little effect on the ICCs (table 2). Although the ICCs were small, the

TABLE 1	Participating centres in the European Community Respiratory Health Survey (ECRHS) I, response rate to stage two and equipment used for lung function measurement		
Centre (country)	Equipment used	Stage two response	Nonsmoking asymptomatic participants with lung function[#]
Antwerp South (BE)	SensorMedics spir.	558 (69.7)	148 (28)
Antwerp City (BE)	SensorMedics spir.	564 (65.1)	96 (16)
Hamburg (DE)	Jaeger pneum.	1252 (38.3)	289 (66)
Erfurt (DE)	Jaeger pneum	731 (67.9)	229 (46)
Barcelona (ES)	Biomedin spir.	392 (73.5)	58 (16)
Galdakao (ES)	Biomedin spir.	486 (84.7)	123 (36)
Albacete (ES)	Biomedin spir.	435 (66.1)	131 (36)
Oviedo (ES)	Biomedin spir.	357 (68.4)	74 (20)
Huelva (ES)	Biomedin spir.	271 (56.7)	82 (25)
Bordeaux (FR)	Vitalograph spir.	544 (18.5)	138 (40)
Grenoble (FR)	Biomedin spir.	473 (40.6)	167 (13)
Montpellier (FR)	Biomedin spir.	456 (12.2)	155 (11)
Paris (FR)	Biomedin spir.	651 (20.9)	182 (17)
Dublin (IE)	Biomedin spir.	454 (75.8)	92 (16)
Pavia (IT)	Biomedin spir.	310 (40.0)	90 (17)
Turin (IT)	Biomedin spir.	244 (47.1)	82 (15)
Verona (IT)	Biomedin spir.	242 (67.9)	129 (29)
Groningen (NL)	Morgan spir.	380 (63.4)	124 (35)
Bergen-op-Zoom (NL)	Morgan spir.	452 (70.8)	126 (32)
Geleen (NL)	Morgan spir.	415 (61.8)	117 (26)
Cambridge (UK)	Biomedin spir.	110 (52.6)	99 (17)
Caerphilly (UK)	Biomedin spir.	111 (71.6)	130 (9)
Ipswich (UK)	Biomedin spir.	448 (65.7)	164 (29)
Norwich (UK)	Biomedin spir.	473 (72.2)	151 (24)
Reykjavik (IS)	SensorMedics spir.	563 (83.8)	167 (30)
Bergen (NO)	SensorMedics spir.	835 (72.2)	268 (45)
Gothenburg (SE)	SensorMedics spir.	682 (88.3)	187 (40)
Umea (SE)	SensorMedics spir.	552 (90.3)	183 (41)
Uppsala (SE)	SensorMedics spir.	622 (87.7)	212 (53)
Basel (CH)	SensorMedics spir.	853 (81.4)	272 (70)
Winnipeg (CA)	Spirotech spir.	603 (57.4)	225 (50)
Vancouver (CA)	Spirotech spir.	510 (40.9)	203 (33)
Hamilton (CA)	Spirotech spir.	503 (36.1)	172 (36)
Montreal (CA)	Spirotech spir.	499 (38.9)	168 (48)
Halifax (CA)	Spirotech spir.	255 (24.5)	68 (20)
Prince Edward Island (CA)	Spirotech spir.	596 (35.2)	175 (19)
Wellington (NZ)	SensorMedics spir.	485 (65.5)	131 (18)
Christchurch (NZ)	Spirotech spir.	458 (64.3)	133 (20)
Hawkes Bay (NZ)	Morgan pneum.	316 (57.6)	77 (8)
Portland (USA)	Spirotech spir.	726 (45.3)	161 (21)
Melbourne (AU)	Fleisch pneum.	669 (40.8)	183 (15)
Tartu (EE)	Jaeger pneum.	431 (70.8)	162 (48)
Total		21503 (48.0)	6323 (1234)

Data are presented as n (%), unless otherwise stated. BE: Belgium; DE: Germany; ES: Spain; FR: France; IE: Ireland; IT: Italy; NL: the Netherlands; UK: United Kingdom; IS: Iceland; NO: Norway; SE: Sweden; CH: Switzerland; CA: Canada; NZ: New Zealand; USA: United States of America; AU: Australia; EE: Estonia; spir.: spirometer; pneum.: pneumotach. [#]: number on parentheses is n for those aged <25 yrs.

differences between centres were highly statistically significant (p-value for heterogeneity <0.0001 in each case). Chance accounted for less than half of the observed variation between centres (FEV1 27.6% in males, 32.6% in females; FVC 42.5% and 26.0%, respectively).

ATS criterion

Mean FEV1, and FVC in males, did not differ significantly between measurements that met the ATS criterion and those that did not. Mean FVC was 0.32 L lower (95% confidence interval (CI) 0.09–0.55) in females for whom the criterion

TABLE 2	Mean and components of variance of forced expiratory volume in one second (FEV1) and forced vital capacity (FVC) in adults aged 20–24 yrs who had never smoked or had asthma, and had not wheezed in the last 12 months

Sex	Subjects n	Measure[#]	Mean L	Average within-centre SD L	Between-centre SD L	Intraclass correlation coefficient[¶]	Total single-determination SD L	Regression coefficient of lung function on height L·m^{-1}
Males	605	FEV1	4.76	0.62	0.21	0.10	0.65	NA
Males	605	FEV1 adjusted to height 1.79 m	4.79	0.49	0.17	0.11	0.52	5.41
Females	629	FEV1	3.51	0.47	0.12	0.06	0.48	NA
Females	629	FEV1 adjusted to height 1.66 m	3.54	0.39	0.10	0.06	0.41	3.98
Males	604	FVC	5.60	0.75	0.21	0.07	0.78	NA
Males	604	FVC adjusted to height 1.79 m	5.63	0.58	0.14	0.06	0.60	6.83
Females	629	FVC	4.00	0.58	0.13	0.05	0.59	NA
Females	629	FVC adjusted to height 1.66 m	4.04	0.47	0.10	0.05	0.49	5.16

NA: not applicable. [#]: not height adjusted unless stated; [¶]: the proportion of the total variation that can be attributed to centre influences; it is calculated as: (between-centre SD)^2/((between-centre SD)^2 + (within-centre SD)^2).

was met. This factor did not explain any of the between-centre variation in either measure for either sex (data not shown).

Relationship of centre means at ages 20–24 yrs with response rate and type of instrument

There was no evidence for a relationship of height-adjusted centre mean FEV1 (p=0.72) or FVC (p=0.57) in males or FVC in females (p=0.17) with overall centre response rate, but there was some evidence of an increase for females with response rate for FEV1 (0.021 L per 10% increase in response rate; 95% CI 0.010–0.041; p=0.040). Mean FEV1 by instrument ranged from 4.33 L (Vitalograph) to 5.14 L (Morgan pneumotach) in males, and from 3.40 L (Fleisch) to 3.72 L (Morgan pneumotach) in females (fig. 1). There were corresponding differences
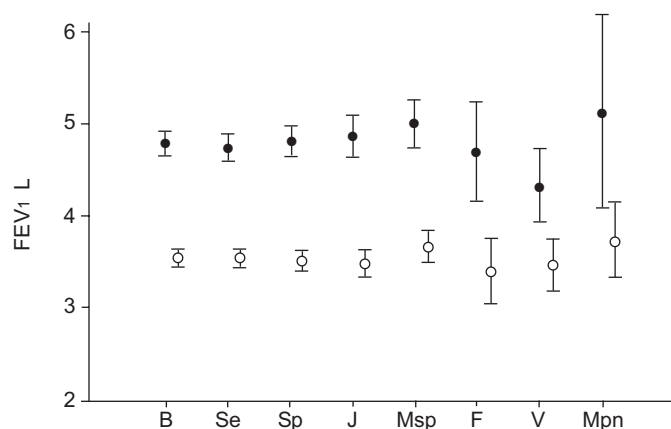

FIGURE 1. Mean forced expiratory volume in one second (FEV1) and 95% confidence interval by instrument type for male (●) and female (○) subjects aged 20–24 yrs. B: Biomedin spirometer; Se: SensorMedics spirometer; Sp: Spirotech spirometer; J: Jaeger pneumotach; Msp: Morgan spirometer; F: Fleisch pneumotach; V: Vitalograph spirometer; Mpn: Morgan pneumotach.

in FVC. The variation in means was statistically significant for FVC in males (p=0.021), accounting for 29% of the centre variation. The differences by instrument type were not statistically significant for FEV1 in males (p=0.06; 20% variation explained), FEV1 in females (p=0.44; <1%) or for FVC in females (p=0.81; 0%). Comparing the three makes of spirometer that were used in more than one country, Biomedin, SensorMedics and Spirotech, there were no significant differences between mean FEV1 or FVC by make with adjustment for country, but country differences adjusted for make were statistically significant for males (FEV1 p=0.0005; FVC p=0.0027). Divided, a priori, into Biomedin, other spirometer and other type of instrument, there were significant differences between the groups in mean FEV1 in males, adjusted for country (p=0.009), but not in the other measures or in means unadjusted for country except for FVC in males (p=0.017).

Figure 2 shows mean FEV1 by country, ranging from 4.18 L (Norway) to 5.02 L (New Zealand) in males, and from 3.23 L (Norway) to 3.68 L (Ireland) in females, with corresponding differences in FVC. The percentage of centre variation explained by country was 52% (p=0.002) and 34% (p=0.035) for FEV1 in males and females, respectively, and 53% (p=0.002) for FVC in males but 0% (p=0.52) for FVC in females. Between-country differences were statistically significant with adjustment for instrument group in males (FEV1 p=0.0006; FVC p=0.0004), but not those in females (FEV1 p=0.073; FVC p=0.348).

Relationships with age and height for subjects aged 25–44 yrs

There was some heterogeneity between centres in the relationship of FEV1 with age (p=0.001 in males; p=0.048 in females) and of FVC to age (p=0.008 in males; p=0.006 in females) adjusted for variation in height. Chance accounted for 55.8%, 71.8%, 62.4% and 60.7% of the observed between-centre
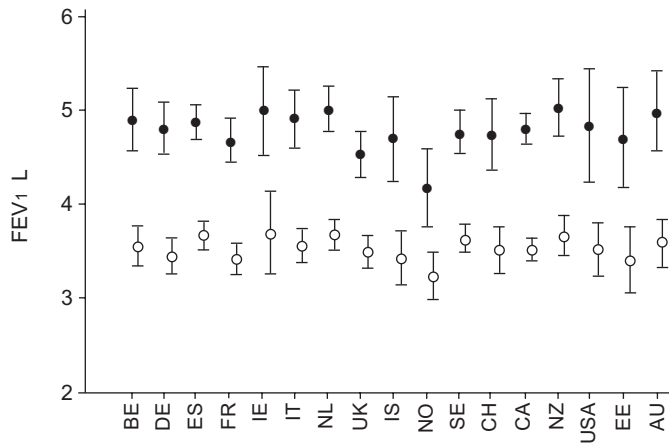
FIGURE 2. Mean forced expiratory volume in one second (FEV1) with 95% confidence intervals by country for male (●) and female (○) subjects aged 20–24 yrs. BE: Belgium; DE: Germany; ES: Spain; FR: France; IE: Ireland; IT: Italy; NL: the Netherlands; UK: United Kingdom; IS: Iceland; NO: Norway; SE: Sweden; CH: Switzerland; CA: Canada; NZ: New Zealand; USA: United States of America; EE: Estonia; AU: Australia.

variation in the relationship with age, respectively. There was heterogeneity in the relationship of FEV1 with height (p=0.009) and of FVC with height (p=0.003) in females, adjusted for age, but not in males (p=0.809 and 0.291, respectively). There was no association between the relationship with age in those >25 yrs and the mean for participants aged 20–24 yrs across centres for either measure in males or females. The estimated relationships with age from subjects aged 25–44 yrs are shown in table 3. The estimated linear decline in FEV1 with age did not differ significantly between males and females (p=0.283), and there was weak evidence for a difference in FVC with age (p=0.046). However, there was some evidence for the relationship of each measure with age being nonlinear in females, with a greater decline in those aged >40 yrs (data not shown). FEV1 and FVC each increased with each centimetre of height by more in males than in females (table 3). Components of variance in table 2 were virtually unchanged on adjustment for nonstatistically significant relationships of height with age within the age range 20–24 yrs.

DISCUSSION
Of the four elements of adult reference equations, *i.e.* mean in early adulthood, relationship with height, decline with age and

variation around the predicted value, it has been shown that variation in the mean values at the youngest mature age may differ substantially between populations. Variation in the other components between populations is likely to have less influence or almost none. Although the current study data were obtained in young adults, and variation in relationships with age may be more important at older ages, mean values at the age of maximum lung function are important components of reference equations.

Although there was evidence for variation between centres in the relationship of lung function with age, more than half of the observed variation could be attributed to chance. For most uses, in diagnosis and patient selection, cross-sectional reference curves are appropriate. Several authors have found differences between cross-sectional associations with age and longitudinal decline with age [17], but not all in a consistent direction [18, 19]. The longitudinal decline with age should represent the true mean decline due to the ageing process, but estimates may be affected by selective participation in multiple surveys, learning effects and healthy survivor effects. Cross-sectional relationships with age will encompass cohort effects as well as healthy survivor effects, but are less affected by participation bias than longitudinal estimates. These different influences may explain the discrepancies between the cross-sectional and longitudinal findings. Clearly, it is desirable to allow for pure cohort effects in reference equations, but part of the decline with age may be due to increasing ill health, and earlier cohorts may have poorer health than later cohorts. Hence, full age adjustment may lead to underdiagnosis of lung disease. Although never-smokers without wheeze or asthma were selected, in common with most studies that reported reference equations, asymptomatic disease could not be rules out. The increasing decline with age, observed in the current females (data not shown) and reported by others [17, 18, 20], may also be due to effects of poorer health.

The current authors chose to compare means in those aged 20–25 yrs since several studies have estimated the decline after that age [8, 11], while others have modelled means from an earlier age [14, 20–23]. Studies that have modelled variation from subjects aged ≤20 yrs do not show a consistent age of maximum lung function. FALASCHETTI *et al.* [14] have found an earlier decline in FEV1 in females than in males but an extended plateau in FVC, GULSVIK *et al.* [20] have shown an apparent maximum lung function in male subjects aged 30 yrs

TABLE 3	Relationship of forced expiratory volume in one second (FEV1) and forced vital capacity (FVC) with height and age in adults aged ≥25 yrs, who had never smoked or had asthma, and had not wheezed in the last 12 months			
Sex	**Subjects n**	**Measure**	**Regression coefficient (95% CI) on age combined across centres# mL yr-1**	**Regression coefficient (95% CI) on height combined across centres# L m-1**
Males	2344	FEV1	-24 (-28– -19)	5.2 (4.9–5.5)
Females	2742	FEV1	-21 (-23– -18)	3.9 (3.6–4.2)
Males	2344	FVC	-17 (-22– -11)	7.3 (6.9–7.7)
Females	2742	FVC	-12 (-16– -8)	5.4 (5.0–5.7)

CI: confidence interval. #: using random-effects meta-analysis.

and stated that curves in males and females were parallel, and LANGHAMMER *et al.* [23] have demonstrated a decline in male subjects from 20 yrs but later in females. These differences may be an artefact of the various forms of equations used to model mean lung function. It is unlikely that the sample sizes available, ~6,000 in the Health Survey for England across the age range of 16–85 yrs and over [14], would be adequate to distinguish between these models.

By far the most difficult issue is what mean value of each measure of lung function in subjects aged 20–25 yrs, or at the age of maximum lung function, should be used in reference equations. Although considered as random effects, the centre differences increased the residual standard deviation, and hence the width of a reference range, by at most 6% (table 2), the differences in means by type of instrument were not negligible (fig. 1). Compared with between-centre variation, differences between the eight types of equipment were not generally statistically significant, but this analysis of 42 mean values does not rule out important real differences. Assuming a normal distribution, and, therefore, using the mean minus 1.64 total standard deviations of FEV1 from table 2 [14, 21], the estimated 5th centiles of FEV1 for males and females of average height aged 20–25 yrs are 3.94 L and 2.87 L, respectively, but range 3.48–4.29L for males and 2.73–3.05 L for females using the minimum and maximum means from figure 1 with within-centre variation from table 2. The ATS [13] and the European Respiratory Society [11] each state that calibration of equipment should achieve readings to within $\pm$50 mL.

Although participation bias could be ruled out as unlikely to be a major cause of the centre variation, true population differences could not be fully separated from instrument variation. The latter comprises spirometer *versus* pneumotach difference, make, model and machine within-model variation, and inconsistent calibration and operation. It is not possible to ascertain whether corrections to body temperature, ambient pressure, saturated with water vapour conditions are comparable between different manufacturers. In so far as the same type of instrument was used in several centres in different countries, there is evidence that there are true population differences and that differences between spirometers may be less important, but neither of these can be quantified from the current study. However, there is evidence that even devices of the same type, used under carefully controlled conditions and calibration, may give differing results [24, 25].

The current results suggest that centre variation was more likely to be due to true population differences. Population differences may be due to genetic differences or to differences in health that are not removed by restricting the data to nonsmokers without wheeze or asthma. In the present data, differences remained after adjusting for variation between countries in height and also for variation in body mass index. As participants in the ECRHS were almost exclusively White, the country differences lead to the conclusion that reference curves cannot be guaranteed to be applicable to a population of the same ethnic group as that from which they were derived, as recently recommended [6]. Conversely, national reference curves cannot be used in epidemiological studies that seek to compare populations

and, if used to select patients with lung function below a given percentage of predicted value in multicentre trials across different countries, may result in heterogeneity in severity of disease in those chosen.

Without a large international study comparing several instruments within each centre, it is impossible to fully separate instrument from true population differences. Studies that wish to compare population values between centres must be prepared to invest in standardised equipment. In order to show whether population differences in lung function represent differences in health, it would be necessary to compare mortality and morbidity between populations in relation to lung function measured in a standardised way. Neither national reference curves nor ones based on the same ethnic group can be guaranteed to give accurate norms of lung health.

REFERENCES

1 Pauwels RA, Buist AS, Calverley PMA, Jenkins CR, Hurd SS. Global strategy for the diagnosis, management, and prevention of chronic obstructive disease. NHLBI/ WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am J Respir Crit Care Med* 2001; 163: 1256–1276.

2 Anthonisen NR, Connett JE, Murray RP, for the Lung Health Study Research Group. Smoking and lung function of lung health study participants after 11 years. *Am J Respir Crit Care Med* 2002; 166: 675–679.

3 Schünemann HJ, Grant BJB, Freudenheim JL, *et al.* The relation of serum levels of antioxidant vitamins C and E, retinol and carotenoids with pulmonary function in the general population. *Am J Respir Crit Care Med* 2001; 163: 1246–1255.

4 Willemse BWM, Postma DS, Timens W, ten Hacken NHT. The impact of smoking cessation on respiratory symptoms, lung function, airway hyperresponsiveness and inflammation. *Eur Respir J* 2004; 23: 464–476.

5 Roca J, Burgos F, Sunyer J, *et al.* References values for forced spirometry. Group of the European Community Respiratory Health Survey. *Eur Respir J* 1998; 11: 1354–1362.

6 Miller MR, Crapo R, Hankinson J, *et al.* General considerations for lung function testing. *Eur Respir J* 2005; 26: 153–161.

7 Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ATS recommendations. *Am Rev Respir Dis* 1981; 123: 659–664.

8 Knudson RJ, Lebowitz MD, Holdberg CJ, Burrows B. Changes in normal maximal expiratory flow-volume curve with growth and aging. *Am Rev Respir Dis* 1983; 127: 725–734.

9 Paoletti P, Pistelli G, Fazzi P, *et al.* Reference values for vital capacity and flow-volume curves from a general population study. *Bull Europ Physiopathol Respir* 1986; 22: 451–459.

10 Castellsague J, Burgos F, Sunyer J, Barbera JA, Roca J. Prediction equations for forced spirometry from European origin populations. Barcelona Collaborative Group on

Reference Values for Pulmonary Function Testing and the Spanish Group of the European Community Respiratory Health Survey. *Respir Med* 1998; 92: 401–407.

11 Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault J-C. Lung volumes and forced ventilatory flows. *Eur Respir J* 1993; 6: Suppl. 16, 5–40.

12 Burney PG, Luczynska C, Chinn S, Jarvis D. The European Community Respiratory Health Survey. *Eur Respir J* 1994; 7: 954–960.

13 American Thoracic Society. Standardisation of spirometry. *Am J Respir Crit Care Med* 1995; 152: 1122.

14 Falaschetti E, Laiho J, Primatesta P, Purdon S. Prediction equations for normal and low lung function from the Health Survey for England. *Eur Respir J* 2004; 23: 456–463.

15 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7: 177–188.

16 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analysis. *BMJ* 2003; 327: 557–560.

17 Vollmer WM, Johnson LR, McCamant LE, Buist AS. Longitudinal *versus* cross-sectional estimation of lung function decline: further insights. *Stat Med* 1988; 7: 685–696.

18 Vollmer WM, Johnson LR, McCamant LE, Buist AS. Methodologic issues in the analysis of lung function data. *J Chronic Dis* 1987; 40: 1013–1023.

19 Ware JH, Dockery DD, Louis TA, Xu X, Ferris BG, Speizer FE. Longitudinal and cross-sectional decline in never smoking adults. *Am J Epidemiol* 1990; 132: 685–700.

20 Gulsvik A, Tosteson T, Bakke P, Humerfelt S, Weiss ST, Speizer FE. Expiratory and inspiratory forced vital capacity and one-second forced volume in asymptomatic never-smokers in Norway. *Clin Physiol* 2001; 21: 648–660.

21 Brändli O, Schindler CH, Künzli N, Keller R, Perruchoud AP. Lung function in healthy never smoking adults: reference values and lower limits of normal of a Swiss population. *Thorax* 1996; 51: 277–283.

22 Gore CJ, Crockett AJ, Pederson DG, Booth ML, Bauman A, Owen N. Spirometric standards for healthy adult lifetime nonsmokers in Australia. *Eur Respir J* 1995; 8: 773–782.

23 Langhammer A, Johnsen R, Gulsvik A, Holmen TL, Bjermer L. Forced spirometry reference values for Norwegian adults: the Bronchial Obstruction in Nord-Trøndelag study. *Eur Respir J* 2001; 18: 770–779.

24 Künzli N, Ackermann-Liebrich U, Keller R, Perruchoud AP, Schindler C. Variability of FVC and FEV₁ due to technician, team, device and subject in an eight centre study. *Eur Respir J* 1995; 8: 371–376.

25 Künzli N, Kuna-Dibbert B, Keidel D, *et al.* Longitudinal validity of spirometers: a challenge in cohort studies. *Swiss Med Wkly* 2005; 135: 503–508.